Minds & Machines (2008) 18:357–372
DOI 10.1007/s11023-008-9105-7

Embodying Emotions: What Emotion Theorists Can Learn from Simulations of Emotions

Matthew P. Spackman · David Miller

Received: 16 April 2007 / Accepted: 16 June 2008 / Published online: 2 July 2008

Abstract Cognitively-oriented theories have dominated the recent history of the study of emotion. However, critics of this perspective suggest the role of the body in the experience of emotion is largely ignored by cognitive theorists. As an alternative to the cognitive perspective, critics are increasingly pointing to William James' theory, which emphasized somatic aspects of emotions. This emerging emphasis on the embodiment of emotions is shared by those in the field of AI attempting to model human emotions. Behavior-based agents in AI are attempts to model the role the body might play in the experiencing of emotions. Progress in creating such behavior-based models that function in their environments has been slow, suggesting some potential problems with Jamesian alternatives to cognitive perspectives of emotions. Heidegger's and Merleau-Ponty's conceptions of embodiment are suggested as alternatives to James' and as means for addressing the shortcomings of the cognitive perspective.

Keywords Appraisal theory · Artificial intelligence · Embodiment · Emotion · Heidegger

Since the publication of Arnold's landmark text (Arnold 1960), cognitively-oriented theories of emotion have dominated the study of emotion (see Oatley 2004; Shields and Kappas 2006; Solomon 1993b, for discussions). This dominance is found in both psychological approaches to emotions (see, for example, Lazarus 1991; Scherer et al. 2001; Smith and Lazarus 1993), as well as machine models of emotions in artificial intelligence and the cognitive sciences (as in Minsky 1988; Picard 1997; Sloman 1987). However, within both of these fields, there is a growing

M. P. Spackman (✉) · D. Miller
Department of Psychology, Brigham Young University, 1001 Spencer W. Kimball Tower, Provo, UT 84602, USA
e-mail: matt_spackman@byu.edu

movement toward a more Jamesian perspective of emotions (see Damasio 1994; Griffiths 1997; Prinz 2004b) or what cognitive scientists refer to as an embodied–embedded approach to emotions (see Clark 1997; DeLancey 2002; Dreyfus 1992; Gallagher 2005; Varela et al. 1993; Wheeler 2005). Though the Jamesian and embodied–embedded movements differ in respects important to this paper, they share in common their opposition to the cognitivist assertion that emotions are reducible to cognitive appraisal processes.

Foundations of Cognitive Approaches

Prinz (2004a, b) suggests cognitive theories of emotion in philosophy and psychology typically share at least two foundational assumptions: the conceptualization hypothesis and the disembodiment hypothesis. The conceptual hypothesis is the suggestion that emotions require and can be explained in terms of concepts, mental representations that may involve thoughts, beliefs, attitudes, or appraisals. The disembodiment hypothesis claims emotions are separable from, and may not require, any somatic concomitants. Both of these hypotheses are discussed below.

The conceptual hypothesis claims mental representations such as beliefs or appraisals are necessary to emotions. As Lyons (1985) explains, "A cognitivist theory of emotion is one that makes some aspect of thought, usually a belief, central to the concept of emotion, and at least in some cognitive theories, essential to distinguishing different emotions from one another" (p. 21). The representations cognitive theorists suggest comprise emotions may be described as propositional attitudes. The mental state for any given emotion is composed of both a proposition, which expresses a state of affairs or declares something to be the case, and some attitude toward that proposition. A proposition-attitude pair is typically joined by a "that-clause." Thus, the subject's attitude (Sam fears, believes, hopes, doubts, etc.) will be followed by "that" and some proposition (gas prices will increase). In this example, the state of affairs is the propositional object of Sam's mental state. According to Griffiths (1997), this approach to emotion is assumed to be a logical result of the uncontroversial idea that emotions involve mental states, which represent states of the external world.

While philosophers endorsing a cognitivist view typically define emotions in terms of what they refer to as their propositional attitudes, psychologists tend to discuss emotions as the result of a set of appraisals. Prinz suggests however, that this difference in terminology is superficial. According to appraisal theories of emotion (see, for example, Lazarus 1991), when a person has an emotion, they are really forming a judgment, or group of judgments. Such judgments are mental representations of the relationship between an organism and its environment and of how that relationship affects the organism's well being. This definition is also applicable to propositional attitude theories in philosophy, given that such theories generally involve the assumption that propositional attitudes derive from representations of one's well being. In addition to a commitment to conceptual judgments, both psychologists and philosophers also assume that the appraisals entailed in a particular emotion may distinguish that emotion from another. Thus, as Prinz

(2004b) suggests, "What distinguishes two emotions is a matter of the themes they designate: danger, offense, infidelity, self-achievement, etc." (p. 25).

Whereas most propositional attitude theorists simply claim that the propositional component is necessary to emotion, some further assert that the propositional component itself is sufficient for the occurrence and explanation of emotion (for examples, see Nussbaum 1986, 1994; Scherer et al. 2001; Solomon 1988). Thus, the conceptual element in and of itself could be fully considered an emotion. Solomon (1984, 1988), for example, clearly makes this reductive claim (see also discussion of Solomon in Mazis 1993; Prinz 2004b).

The second unifying assumption of cognitive theories, according to Prinz (2004b), is the disembodiment hypothesis. This is the claim that the cognitive elements of an emotion are distinct from any bodily change that may accompany an emotion. As with Solomon (1993a), several theorists argue that emotions can occur without any somatic component. Others, however, suggest that somatic components are necessary, but require nonsomatic cognitive states as well (for discussion of this issue, see Solomon 2004b). Cognitive theories of emotion share in common a clear theoretical distinction between the somatic/bodily aspects of emotions and their cognitive appraisals and representations. Such theories typically either attribute primacy over somatic changes or bodily states to the cognitive components of emotion, or suggest the sufficiency of the cognitive components in the production of emotions, thereby making propositional judgments disembodied.

Critiques of Cognitive Approaches

Though cognitive approaches to emotion remain dominant within psychology, a number of significant critiques of the approach have been offered. Several arguments focus specifically on the assumptions and implications of the conceptualization hypothesis, as described by Prinz (2004a, b). A number of theorists (both psychologists as well as philosophers) have also argued from a Jamesian perspective that emotions can be experienced in the absence of cognitive appraisals (Damasio 1994; DeLancey 2002; Griffiths 1997; Parkinson 1995; Prinz 2004b). Such arguments counter the claims of the disembodiment hypothesis of cognitivist theories, suggesting that, though our understandings of our surroundings may be important to the experiences of many emotions, as cognitivists would hold, there are also times when emotions occur in the absence of such appraisals.

Griffiths (1997) for example, offers arguments challenging the necessity of cognitive appraisal for emotion. The first of these is the existence of objectless emotions. Experiences of emotional states such as depression, elation, and anxiety are generally thought to be possible in the absence of intentional objects, and thus to potentially involve no propositional attitudes. The generalized feeling of being depressed does not necessarily involve being depressed *about* any specific thing, and so may not require a set of conceptual beliefs. Given the cognitivist's argument that emotions are about something, Griffiths suggests the only way to make sense of such instances, from within the cognitivist framework, is to say that the object of a state of depression is "things generally" (p. 28).

A second example that challenges the conceptualization hypothesis is the occurrence of "reflex" emotions. As Griffiths (1997) argues, a person may hold the conviction that worms are completely harmless and yet simultaneously experience emotions of disgust or fear in the presence of worms, which directly contradicts the judgment that worms are harmless (see also Calhoun 2004, for discussion of this issue). Whereas a cognitive theory of emotion would explain such emotional responses in terms of a specific appraisal, it may be argued such appraisals never occur, even though the emotional response was experienced.

An additional challenge to the cognitivist position is constituted by the idea of unemotional evaluations. For example, many smokers judge smoking to be dangerous and yet smoke without experiencing any fear. To make sense of instances of unemotional evaluations would require some explanation for why only some evaluative judgments are emotions (see Solomon 2004a, for a response to this critique). Unemotional evaluations also demonstrate an additional area of difficulty for cognitive theories of emotion: judgments may well underdetermine emotions. An example of such under determination offered by Lyons (1985) is the judgment that Ashkenazy is a fine pianist. This judgment may give rise to any of several emotions—envy, admiration, or even no emotion at all. The occurrence of such a judgment does not, however, determine which of the possible emotional reactions that may arise from the judgment actually occurs (if any). Examples such as this suggest that the judgments associated with a particular emotion do not fully account for or determine that emotion, and that emotions may not be distinguishable by their cognitive contents.

One final problem that demonstrates the difficulty of coupling emotion with judgment is the experiencing of emotional responses to imagination. Several philosophers have argued that people can experience genuine emotions by simply imagining corresponding objects or situations. In such circumstances, the subject does not possess the types of beliefs the propositional attitude approach claims are necessary to account for emotion (for discussion of this issue, see Robinson 2005).

Whereas these arguments directly challenge the conceptualization hypothesis underlying cognitivist theories, a number of similar arguments also pose challenges for the disembodiment hypothesis. A central feature of the disembodiment hypothesis is the distinction made between the mental or cognitive components of emotion and the somatic components. Critics of the cognitive approach argue that cognitive theorists problematically neglect the physiological aspects of emotion. For example, research suggests emotional responses may be distinguished and characterized by at least four kinds of physiological change: facial expressions, musculoskeletal changes (flinching, etc.), expressive vocal changes, and autonomic nervous system changes such as adrenaline release and change of heart rate (Griffiths 1997). Despite the examination of these physiological events in researching emotion, and even in determining whether or not an emotion has occurred, propositional attitude theories typically suggest emotions may occur in the absence of any such physiological changes. Theorists such as Zajonc (Zajonc 1980, 1984; Zajonc et al. 1989), however, have challenged the cognitivists' claim of the primacy of cognition over somatic aspects of emotion afforded by the disembodiment hypothesis.

Zajonc (1980, 1984) has also claimed that emotion can occur without cognitive accompaniment at all, famously defending the "primacy of affect," the suggestion that an emotion can occur prior to and independent of related cognitive states. Zajonc offers several lines of argument in support of this thesis. First, he suggests emotional expressions and behaviors are observed in human infants and nonhuman animals without reasons for assuming a necessary cognitive component. Emotion and cognition are also argued to involve separate neuroanatomical structures, which further supports the lack of a necessary cognitive component to emotion. In addition, there is evidence for a direct pathway from the retina to the hippocampus, which Zajonc believes might mediate between visual perception and fear response without any cognitive mediation in the neocortex (see also LeDoux 1993, 1995, 1998). Thus, encountering a coiled, snake-like object can produce fear prior to information arriving at the neocortex and permitting the object to be recognized.

As a final argument supporting the primacy of emotion, Zajonc claims that emotional states can be induced without any prior mental states. In other words, emotions can be induced by more direct physical means: drugs, hormones, electrical stimulation, even changes in facial expression or body posture can result in changes in emotion (for representative research, see Laird 1974; Laird and Bresler 1990).

A potential solution to many of the problems facing cognitivism offered by critics such as Prinz and others (Damasio 1994; DeLancey 2002; Griffiths 1997; Parkinson 1995; Prinz 2004b) is to suggest a necessary role in the formation of emotion for the body, as James (1884) famously suggested long before the cognitive movement. A Jamesian theory defines emotion as our perception of physiological excitement, which is based on our perception of objects or events in our environment. In his early essay on emotion, James stated: "My thesis…is that the bodily changes follow directly the perception of the exciting fact, and that our feeling of the same changes as they occur IS the emotion" (James 1884, pp. 189–190). In this approach to emotion, bodily states are given primacy in determining emotions. Thus, an underlying difference between Jamesian and cognitively-oriented theories of emotion is that Jamesian theorists suggest the body is necessary to the experience of emotions, whereas cognitivists would claim that emotion can be experienced in the absence of any sort of feeling state (see Ekman and Davidson 1994, for discussion on this issue). Additionally, Jamesian theorists would suggest that emotions may occur in the absence of cognitive appraisals. The necessary role of the body in emotions suggested by Jamesian theorists is an idea that has also been receiving increasing attention from cognitive scientists and others working in the field of AI.

AI and Embodiment

The question of the necessity of bodily involvement in the experience of emotions has been important not only to current debates on the nature of emotion in psychology and philosophy, but also to contemporary attempts at emotion simulation in the field of artificial intelligence as well. Since its beginnings, the field of AI has served as a primary area of testing and application for theories in

psychology and cognitive science (Wheeler 2005). AI often empirically illustrates theoretical problems and limitations at the conceptual level (see Anderson 1993, 1996, for the ACT-R modeling language designed to do just this sort of theory testing). As computational theories of information processing models are employed in the design of intelligent robotic systems, the testing and development of those systems contributes to our understanding of the nature and problems of human cognition. According to Brooks (1999), AI itself attempts both to build useful intelligent systems, and contribute to the general understanding of human functioning. Thus, while the construction of such systems may be based on theories from psychological fields, their application and testing contributes to our understanding of human functioning by permitting empirical evaluation of theory.

The field of artificial intelligence is generally considered to have two primary goals. First is the re-creation of intelligence or intelligent behaviors. Second is the aim of creating the level of independence or autonomy demonstrated by humans and other animals. Thus, autonomy is argued to be a central problem of AI. In its infancy, the field of AI experienced a nearly unbounded optimism for the possibility of creating such machines. However, with the unfolding of significant obstacles, and the appreciation of the magnitude and difficulty of this goal, such hopes were somewhat diminished before being reframed under new conceptual approaches (Brooks 1999; Dreyfus 1992; Varela et al. 1993).

Though the goals and optimism of early AI were fairly ambitious (replicating human intelligence and autonomy), the approaches were usually more focused and modest. Much of the earliest work in AI focused on specific tasks such as game playing, language translation, and basic problem solving. This stage in the development of AI was largely based on the cognitive simulation approach, in which cognitive heuristics employed by humans were uncovered and then formalized into specific rules for machine programming (Brooks 1999; see also Kahneman and Tversky 2000). For example, early chess playing programs were constructed in which decisions were made on the basis of heuristics assumed to be employed by human chess masters (see Newell et al. 1958). Such programs calculate possible moves, given the rules and current game state, and then selected from among this wide range of possible moves by employing heuristic evaluations of likely game-winning strategies (i.e., "avoid moves that sacrifice the queen"). While programming attempts such as these were quite successful in simpler games, chess and other more complicated decision making environments began to expose fatal limitations inherent to this approach to intelligent decision making (see Dreyfus 1992; Varela et al. 1993).

Though some researchers have made modest attempts at modeling traditional conceptual approaches to emotion, the field of AI has largely ignored emotion (Delancey 2002; though, see Minsky 1988, 2006, for a notable exception here). A number of programs have been designed with the goal of recreating specific aspects of emotion from within the cognitivist framework common to other projects in AI. PARRY, for example, was created in 1972 to interact in a way that seemed to reflect affective responses of a person with paranoid schizophrenia (Colby 1974). The affective reasoner (AR) program was similarly designed to respond to speech with

various multimedia reactions, conveying a basic sense of emotion to its users (Elliott 1994).

Whereas programs such as these may differ in their specific goals and approaches, they reflect underlying cognitivist assumptions about emotion. According to Delancey (2002), these systems are modeled on symbolic computational functionalism. This foundational approach to much of AI design can be considered a manifestation of cognitive theories in that it assumes a hierarchical model of the mind, wherein simple concepts are built level-upon-level, into a central representation of a problem. Accordingly, minds are thought of as information processors which function through the manipulation of discrete symbols (see also Gallagher 2005; Varela et al. 1993). Out of these abstracted symbols, a representational model of the world is produced from which intelligent decision making is theorized to occur. This approach is congruent with the conceptualization hypothesis underlying cognitive approaches to emotion in which affective states are assumed to be reducible to propositional attitudes (see Minsky 2006, as an example of such an approach). These propositional attitudes are themselves composed of discrete conceptual units or symbols, which are processed in a hierarchical fashion, to achieve intelligent behavior.

Brooks (1999) suggests a rethinking of the general problem of organizing intelligence began around 1984. At this time, autonomous mobile robots were designed with the goal of functional operation in environments in which obstacles were not known ahead of time; thus the agents were expected to interact with dynamic, rather than static environments. A key advance that made this goal achievable was a new approach to the underlying architecture. Traditional mobile robots used a hierarchical information processing approach in which input from sensors proceeds from lower level perceptual processes to higher level executive processes before actuating motors for movement. As previously suggested, this layered approach is representative of cognitivist theories of mind. Newer approaches broke from this model by employing behavior-based architecture. Rather than a hierarchical processing of information from sensation to action, robots employing behavior-based architecture approached multiple behavioral goals simultaneously, each of which directly linked perception with action. While each layer may imply others, specific behaviors do not depend on any lower level processes for successful completion. Ultimately, if successfully implemented, such an architecture would mean that no central model or symbolic representation of the world is necessary to achieving intelligent behavior (see also Varela et al. 1993).

Central to this change in design were basic assumptions that challenged representational claims in cognitive theories of mind. For example, it was theorized that most behavior humans engage does not consist of problem solving, which requires symbolic representations of the world. Rather, agents are said to interact with dynamic environments using separate processes that do not make use of central, symbolic data structures. This argument gained support from the discovery of subcortical pathways that allow for emotional responses, such as fear, without requiring cognitive activity in the neocortex (c.f., LeDoux 1993, 1995, 1998).

Another difference underlying this new approach to AI was the need for robots to interact with changing, unpredictable environments, as opposed to the

oversimplified, static ones in which earlier agents were designed to interact. While more traditional agents were essentially cognitivist in their design, more contemporary attempts have embraced an embodied–embedded approach in which dynamic representational systems incorporate embodied appraisals (as opposed to cognitive appraisals) into their programming architecture (see DeLancey 2002; Wheeler 2005). Indeed, Wheeler claims that the cognitive sciences are undergoing a period of dramatic reconstruction in this respect. Newer agents differ from those modeled on cognitive theories in that their representational architectures are not static, as cognitivist systems would be, and newer systems incorporate feedback from their "bodily" systems into their emotional architecture.

Embodied robots are theorized in terms of actions that form a dynamic whole with their world, and a direct experiencing of that world, through behavior-based architecture rather than traditional information processing. Embodiment also signifies that agents' behaviors themselves impact their sensations, just as other embodied agents' actions are thought to. Embodied machines are conceived of as agents directly situated in the world, rather than abstracted from it or requiring symbolic representations of the world to produce intelligent, autonomous behavior. From this perspective, (and given behavior-based design) the environment has a direct effect on behavior, as opposed to traditional forms of information processing in which sensory information is processed layer upon layer toward a central representation of the external world.

This shift away from cognitively-oriented agents and toward more embodied simulators within the AI community has resulted from an increased desire within the community to develop agents that can function in more realistic environments and a growing body of evidence indicating that more orthodox simulators cannot manage such environments (Dreyfus 1992; Fodor 1983; Wheeler 2005). A basic problem with the traditional models is what has been called the frame problem (see McCarthy and Hayes 1969). Agents may be programmed to assess the significance of objects to their goals in limited contexts and, within such contexts, function relatively well. However, as these contexts become increasingly complex, agents perform increasingly poorly. This is because rules dictating the agent's actions are representational in nature. The significance afforded objects in the world is dictated by an algorithmic-based programming architecture. As the number and variety of the objects encountered increases, the architecture dictating the agent's actions begins to collapse under the weight of combinatorial explosion. This is, in part, due to the exponential growth of possible decisions (or pathways) an agent must consider when making a judgment. Given limited processing capabilities, such programs are argued to be incapable of truly intelligent decision making when such expansive possibilities for action are introduced.

However, this argument of combinatorial explosion, though itself a powerful critique of representational systems, may, in theory at least, be answerable simply by suggesting that the resolution lies in developing increasingly complex and powerful representational systems. The more damning critique of representational systems is one of infinite regress. Any representational system must have context-independent algorithms for relevance decisions. These algorithms are necessary to establishing the importance of objects in the environment to the agent's

programmed goals. The algorithms themselves, however, require their own context-independent algorithms, and so on, into infinity. Dreyfus' example of this infinite regress problem follows.

> [T]o pick out two dots in a picture as eyes one must have already recognized the context as a face. To recognize this context as a face one must have distinguished *its* relevant features such as shape and hair from the shadows and highlights, and these, in turn, can be picked out as relevant only in a broader context, for example, a domestic situation in which the program can expect to find faces. This context too will have to be recognized by its relevant features… (1992, p. 288, emphasis in the original)

It is this problem of infinite regress of representational architectures that has lead many in the AI community to suggest that cognitively-oriented architectures such as those suggested by cognitive theories of emotion are insufficient for simulations of psychological states such as emotions. Increasingly, theorists have suggested that embodied–embedded architectures may address the shortcomings of representational states (see, for example, Brooks 1999). The basic problem of representational states is, as shown above, that they require context-independent rules for functioning within environments. Remove the need for such rules, and the problem of regress disappears.

Heideggerian Agents

In an attempt to address the problem of infinite regress and to offer a theoretical grounding to the concept of embodied agents, some have suggested that a Heideggerian conception of embodiment may be a useful model for the development of artificial agents (see Dreyfus 1992; Wheeler 2005; Winograd and Flores 1986, for discussions). Martin Heidegger, the German phenomenological/existential philosopher has become increasingly influential in philosophical schools and his own work on the question of embodiment, as well as that of his followers such as Maurice Merleau-Ponty, has influenced such thinkers within the cognitive sciences as Dreyfus (1992), Gallagher (2005), Wheeler (2005) and others (e.g., Varela et al. 1993; Winograd and Flores 1986).

At its most basic level, Heidegger's (1949, 1996) conception of embodiment (what he calls being-in-the-world) rejects what may be considered the Cartesian subject–object split. That is, Heidegger suggests that the traditional, Cartesian philosophical view of perception involves the agent (the subject) engaging the object unidirectionally. The object is passively acted upon and the act of perception involves the agent representing the object (Cataldi 1993; Gallagher 2005; Mazis 1993; Varela et al. 1993; Wheeler 2005). Heidegger's conception of embodiment rejects this subject–object split, suggesting that the essential divide between subject and object entailed in our modern philosophical perspective does not acknowledge the manner in which subjects and objects mutually constitute one another. That is, the act of perception does not involve the subject representing the object as something in the world, but rather involves the subject constituting the object as

something all the while the object is simultaneously constituting the subject as a particular agent.

Though Heidegger suggests the fact that perceivers are embodied—that they experience their worlds as beings with particular bodily engagements—is essential to our understanding of how we experience our worlds, it is left to Merleau-Ponty to more fully develop the phenomenological conception of embodiment (see especially Merleau-Ponty 1962, 1963). Merleau-Ponty, like Heidegger, rejects the Cartesian subject–object split, suggesting that such a divide compartmentalizes or brackets the subject apart from the world in which the object resides. The traditional conception never allows the subject to directly experience the object, only its representation. Likewise, the object never contacts the subject as she is, in a very Cartesian sense, never truly engaged in the physical world of the object. Merleau-Ponty's conception of embodiment bridges this subject–object divide by reminding us that our experiences of the world are always and only as embodied beings. That is, our bodies both permit and also limit the manner in which we engage the world—they are our engagements with it.

As theoretical grounds on which embodied agents might be constructed, Heidegger's or Merleau-Ponty's conceptions of embodiment would suggest the construction of agents that incorporate an embodiment that is context-dependent. In other words, the agent's behavior-generating architecture would incorporate the agent's embodied interaction with its environment without employing representational states. To this point, the push within the field of AI for Heideggerian or Merleau-Pontian agents has seen relatively little success.[1] Wheeler (2005) identifies a number of other programs of a Heideggerian nature that have made progress, but points out that difficulties remain. One such program, an agent designed by Franceschini et al. (1992), was capable of navigating toward a light source using a visual system accounting for its own bodily motion. A primary difference between this and more traditional agents was that the representations used were context dependent and dynamic. Rather than a central internal representation of the world, this agent used more "action-oriented" representations.

Despite the fact that he explicitly states he does not consider his program of research to be Heideggerian, Brooks' (1999) work on embodied/embedded agents may be seen to share some of the goals of what might be called a Heideggerian program in AI. Brooks has for years attempted to construct agents that are contextualized and which operate non-representationally. However, despite the fact that many of Brooks' agents do not represent their environments at a central level, these agents still represent their environments at a local level and are in that sense not fully Hedeggerian.

Though the development of agents that account for their own bodily motions is a step in the direction of truly embodied/embedded agents, there remain those that are doubtful that programming architectures can ever fully mirror human embodiment in a Heideggerian or Merleau-Pontian sense (see especially Dreyfus 1992). Such

[1] It should be noted that Heidegger's philosophy has historically been more influential to the cognitive sciences than has Merleau-Ponty's. This is primarily due to the early influence of Dreyfus's (1992) Heideggerian critique of AI. Merleau-Ponty's philosophy has, however, received more attention as of late.

critics suggest that even agents who represent their own bodily states still maintain the Cartesian subject–object split as these representations are still just that, representations of embodiment as opposed to simple embodiment itself. Until agents are able to simply be-in-the-world without representing the world, critics suggest the problems of infinite regress and the other shortcomings of AI discussed above will remain.

It may well be asked how we would know if an agent operating successfully without representation of its environment actually experienced an emotion. This of course raised the specter of the question of other minds (see Nagel 1974). Space does not permit a full discussion here. However, see Spackman (2004) for discussion of a Turing test of emotions.

Embodied Emotions

The difficulties still faced by those attempting to construct embodied agents not dependent upon representational states offers an important lesson to emotion theorists. It may be seen that the currently dominant cognitive theory of emotion is a classical Cartesian model of emotions. Cognitive theories suggest persons' (subjects') evaluations of their environments (their representations of objects) determine their emotions. This conception of emotion is Cartesian in the sense that persons never quite engage the objects of their worlds directly, only by way of their (subjective) appraisals of them. In addition, as has been discussed above, cognitive theorists do not suggest a necessary role for the body in the experience of emotion. In fact, appraisals may just as well be disembodied as the evaluations resulting in emotions are in no sense tied to or dependent upon the physical form of the emoter (for more on the disembodiment of appraisal theories, see Cataldi 1993; Mazis 1993). There is, however, as discussed above, a movement within psychology away from appraisal or cognitively-oriented theories of emotion and toward more embodied or Jamesian approaches to emotions. The failures of context-independent representational architectures in AI point out well the difficulties of the cognitively-oriented theories on which they were based. However, the difficulties still remaining for researchers in the field of AI should offer to those embracing a more Jamesian approach to emotion a cautionary note.

In its original incarnation, the Jamesian approach to emotion suggested emotions result from recognition of states of the viscera (James 1884). Contemporary theories of emotion inspired by James allow the viscera may be important to emotions, but often focus attention on the findings of neuroscientists with respect to areas of the brain active during emotions. Damasio's (1994, 1999) somatic marker theory is certainly an example of such a theory. According to this theory, the decision making process is made more efficient by somatic states that "mark" any possibility under consideration with positive or negative feelings. The gut reaction that may accompany a thought is the result of a somatic marker: a quick, often unconscious association of cost/benefit with an idea. By drawing attention to outcomes through somatic states, certain possibilities are essentially highlighted for the actor. The formation of somatic markers is theorized to take place in the ventromedial area of

the prefrontal cortex. Sensory information is sent to the limbic system, which produces a feeling state and sends signals to other areas of the brain, resulting in corresponding bodily reactions. This overall reaction state is compared with memories of similar instances and a neural network association is formed.

Contemporary theories such as Damasio's somatic marker theory might more appropriately be called *neo*-Jamesian in the sense that they deny the cognitivists' suggestion that appraisal is necessary to emotion, but seem to suggest that the embodiment of emotions need not extend below the brain stem. That is, though Damasio's theory does incorporate bodily feedback, emotions are still conceived of as brain events involving representational states (see Gallagher 2005, for a similar critique). In other words, though the role of the body is acknowledged in the form of somatic markers, the body is important to the formation of emotions only insofar as it is represented in the brain. It may be seen that Damasio's, as well as other neo-Jamesian approaches, are similar to attempts to create Heideggerian-inspired agents in AI. The body is represented in the brain just as it is represented by artificial agents. The neo-Jamesians have still not, however, fully embodied emotions, just as current artificial agents are not fully embodied. In other words, like current Heideggerian agents, neo-Jamesian theories allow for involvement of the body, but may be argued to be more Cartesian than Heideggerian with respect to their specific conception of embodiment. In keeping with the disembodiment hypothesis shown above to be an assumption of cognitive theory, mental and somatic aspects of emotion are considered distinct from one another, an assumption of Cartesian dualism that pervades psychology as a whole (Wheeler 2005).

It is with respect to the neo-Jamesians' focus upon neuronal embodiment that the cautionary note from the field of AI applies. The degree to which the neo-Jamesian is willing to suggest that emotions reside within the brain, and not to extend below the brain stem, is also the degree to which emotions thus conceived are, in fact, products of representational states. That is, neurologically-centered emotions are representations of interactions of agents with the environment, and therefore subject to the criticisms of computational explosion and infinite regress discussed above (see Spackman (2004), for a critique of representational states in emotion drawing on the work of Godel). However, if the neo-Jamesian is willing to allow that emotions are truly embodied in the Heideggerian or Merleau-Pontian sense, he or she may avoid the difficulties of representational states.[2]

We are aware of no emotion theories in psychology embracing a Heideggerian or Merleau-Pontian embodied approach. There are, however, important works by philosophers describing an embodied approach to emotion inspired by the works of Heidegger and Merleau-Ponty. These include works by Cataldi (1993) and Mazis (1993). Both of these authors stress Merleau-Ponty's extension of Heidegger's conception of embodiment. An embodied approach to emotion in the tradition of Heidegger or Merleau-Ponty would, as has been discussed above, include a rejection of the traditional subject–object split. This would include the rejection of the cognitive theorists' conception of the subject experiencing his or her context

[2] I am not suggesting there are no benefits to our understanding gained from AI models employing representational states. However, the limitations of such models must be remembered.

(the object) indirectly via representations or appraisals of that environment. For Merleau-Ponty, one experiences the world as a sort of coming-together or moving-toward, where the actor and her environment mutually affect or constitute one another.

Mazis (1993) captures this concept of the mutual constitution of actors and environments in his conception of *e-motion*. He returns to the etymological roots of the term emotion, which suggest a form of movement (see also Averill 1974; Dixon 2003). Movement entails the assumption of some*thing* moving in space. That thing must be in some way embodied and it must exist some*where*. Emotions are then, the coming together of an embodied being with its environment. This coming together may only be comprehended as an existence of mutually constitutive entities with particular embodiments. The embodiment of the emoter is, then, essential to the emotion itself; it is not a secondary aspect of a subject's appraisal of his environment.

In this respect, it is important to note that an embodied approach to emotion in the tradition of Heidegger or Merleau-Ponty differs greatly from what Prinz calls his embodied appraisal approach (2004b). Prinz's (2004a) embodied appraisals are appraisals of such concerns as "organism/environment relations that bear on well-being" (p. 53) that include perceptions of bodily states. These embodied appraisals maintain a Cartesian subject–object split as the body constitutes an object represented at the level of the mind or brain (the subject). Merleau-Ponty would reject these embodied appraisals as his conception of embodiment entails no subject acting on an object, but a coming together of the two (see Mazis 1993). Cataldi's (1993) example of awe describes well Merleau-Ponty's conception of the mutual constitution of subjects and objects.

> When I am moved into feeling in awe of a sunset's beauty, I am simultaneously being grasped, being felt by its beauty and grasping-feeling this beauty back through feeling in awe of it. I apprehend that I am being apprehended (by the "breathtakingly" beautiful sunset) and that, in being apprehended, I am apprehending emotionally (becoming awe inspired)—as I feel myself becoming more and more *en-grossed* in the beauty of its colors, more and more appreciative of its evocatively "secreting" (or "leaking") its significance "there" through me "here," and more and more aware that an emotional significance, that of awe, is unfolding "here" and that through the unfolding of this emotional significance I am more "open" to the beauty of the sunset. (pp. 119–120, italics in original)

Conclusion

A conception of emotions as embodied in the tradition of Heidegger or Merleau-Ponty would embrace the critique of appraisal theorists offered by the neo-Jamesians: The body must be taken into account in the understanding of emotions. Despite the claims of appraisal theorists to the contrary, cognitive representations of environmental conditions alone are not sufficient grounds for the experience of

emotion. However, simply including representations of bodily or somatic states as objects to also be represented to or by the mind, as the neo-Jamesians suggest, does not address the shortcomings of the cognitive approach. Difficulties experienced by the cognitive science and AI community in developing independent artificial agents relying upon representational programming architectures attest to the need for true embodiment to account for agents' abilities to engage their environments. Embodiment from a Merleau-Pontian perspective suggests the body and the environment mutually constitute one another. This conception of embodiment suggests that emotions are the result of the coming together of our embodied selves with our environments. Environment and body are not represented to or by the mind. Rather, emotions are the result of a moving-toward the world.

This being said, it may well be asked how an understanding of emotions as embodied constitutes an advance in our understanding of the phenomenon. First, an embodied approach to emotion better explains the immediacy of the experience of emotions. By "immediacy," we do not only mean the phenomenological experience of emotions as overwhelming us or coming on us suddenly, though these sorts of experiences are important manifestations of emotions the cognitive and Jamesian theories do not account for well. Emotions may also be said to be immediate in the sense that they are unmediated. That is, they are enacted immediately (in both senses of the word) in the world of our experiences and not by way of representations to the brain. Both cognitivist and neo-Jamesian accounts fall short in this respect as they cannot account for the immediacy of emotions.

Both the cognitivist and neo-Jamesian approaches also fall short of an embodied approach in the sense that they cannot fully account for the meaningfulness of emotions. For the cognitivist, emotions either arise from (see, for example, Lazarus 1991) or are (e.g., Solomon 1988) representations of one's environment. The cognitivist might suggest these representations are what emotions mean (i.e., one's anger *is* recognition of a purposeful wrong). However, as has been shown in the field of AI, such representations are subject to the problem of infinite regress ("purposeful" must be defined, terms used to define purposeful must be defined, etc.) and therefore no basis for the meaning of emotions may be found from a cognitivist perspective (see again here Spackman 2004, on Godel). This same inability to find meaning is evident in the neo-Jamesian approach as such theories differ from the cognitive perspective only in that bodily states are also represented. Somatic representation does not, however, yield an escape from infinite regress (see Footnote 2).

Such an escape may only be found if the necessity of representation is removed, as has been suggested—but not achieved—by researchers in the field of AI (see Dreyfus 1992; Varela et al. 1993; Wheeler 2005). A truly embodied approach in the tradition of Heidegger and Merleau-Ponty offers immediacy to emotions by removing the necessity for representation assumed in the cognitivist and neo-Jamesian accounts. That is, emotions as embodied phenomena are engagements with the world directly with the body. Because of the immediacy of our emotions, their meaningfulness exists within them. The meaning of an emotion arises as a part of the emotion as the emotion is enacted. The meaning of an emotion is not dependent upon a representation in the mind of the emoter (which depends upon

another representation, etc.), but it grows, exists, develops with and as a part of the emotion. Only in such immediacy may be found the meaningfulness of our emotions. Only in embodiment may be found such immediacy.

References

Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist, 51*, 355–365.

Arnold, M. (1960). *Emotion and personality* (Vol. 1). New York: Columbia University Press.

Averill, J. A. (1974). An analysis of psychphysiological symbolism and its influence on theories of emotion. *Journal of Theory and Social Behavior, 4*(2), 147–190.

Brooks, R. (1999). *Cambrian intelligence*. Cambridge, MA: MIT Press.

Calhoun, C. (2004). Subjectivity and emotion. In R. Solomon (Ed.), *Thinking about feeling* (pp. 107–121). Oxford: Oxford University Press.

Cataldi, S. (1993). *Emotion, depth, and flesh*. Albany, NY: State University of New York.

Clark, A. (1997). *Being there: Putting brain, body, and world together again*. Cambridge, MA: MIT Press.

Colby, K. (1974). *Artificial paranoia*. New York: Pergamon Press.

Damasio, A. (1994). *Descartes' error: Emotion reason and the human brain*. New York: Avon Books.

Damasio, A. (1999). *The feeling of what happens*. New York: Harcourt Brace.

DeLancey, C. (2002). *Passionate engines*. Oxford: Oxford University Press.

Dixon, T. (2003). *From passions to emotions*. Cambridge: Cambridge University Press.

Dreyfus, H. (1992). *What computers still can't do*. Cambridge, MA: MIT Press.

Ekman, P., & Davidson, R. (Eds.). (1994). *The nature of emotion*. New York: Oxford University Press.

Elliott, C. (1994). *Mulit-media communication with emotion-driven "believable agents"*. Stanford, CA: AAAI.

Fodor, J. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Franceschini, N., Pichon, J., & Blanes, C. (1992). From insect vision to robot vision. *Philosophical Transaction of the Royal Society, 337*, 283–294.

Gallagher, S. (2005). *How the body shapes the mind*. Oxford: Oxford University Press.

Griffiths, P. (1997). *What emotions really are*. Chicago: University of Chicago Press.

Heidegger, M. (1949). *Existence and being* (W. Brock, Trans.). Chicago: Henry Regnery Company.

Heidegger, M. (1996). *Being and time* (J. Stambaugh, Trans.). Albany, NY: State University of New York Press.

James, W. (1884). What is an emotion? *Mind, 9*, 188–205.

Kahneman, D., & Tversky, A. (Eds.). (2000). *Choices, values, and frames*. New York: Russell Sage Foundation.

Laird, J. (1974). Self-attribution of emotion: The effects of expressive behavior on the quality of emotional experience. *Journal of Personality and Social Psychology, 29*, 475–486.

Laird, J., & Bresler, C. (1990). William James and the mechanism of emotional experience. *Personality and Social Psychology Bulletin, 16*, 636–651.

Lazarus, R. (1991). *Emotion and adaptation*. Oxford: Oxford University Press.

LeDoux, J. E. (1993). Cognition versus emotion, again—this time in the brain: A response to Parrott and Schulkin. *Cognition and Emotion, 7*, 61–64.

LeDoux, J. E. (1995). In search of an emotional system in the brain: Leaping from fear to emotion and consciousness. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 1049–1061). Cambridge, MA: MIT Press.

LeDoux, J. E. (1998). Fear and the brain: Where have we been and where are we going? *Biological Psychiatry, 44*, 1229–1238.

Lyons, W. (1985). *Emotions*. Cambridge: Cambridge University Press.

Mazis, G. (1993). *Emotion and embodiment*. New York: Peter Lang.

McCarthy, J., & Hayes, P. (1969). Some philosophical problems from the standpoint of artificial intelligence. In D. Michie & B. Meltzer (Eds.), *Machine intelligence* (Vol. 4). Edinburgh: Edinburgh University Press.

Merleau-Ponty, M. (1962). *Phenomenology of perception* (C. Smith, Trans.). London: Routledge.
Merleau-Ponty, M. (1963). *The structure of behavior* (A. Fisher, Trans.). Boston: Beacon Press.
Minsky, M. (1988). Emotions and the society of mind. In M. Clynes & J. Panksepp (Eds.), *Emotions in psychopathology: Theory and research* (pp. 171–179). New York: Plenum Press.
Minsky, M. (2006). *The emotion machine*. New York: Simon & Schuster.
Nagel, T. (1974). What is it like to be a bat. *Philosophical Review, 83*, 435–450.
Newell, A., Shaw, J., & Simon, H. (1958). Chess playing programs and the problem of complexity. *Journal of Research and Development, 2*, 320–335.
Nussbaum, M. C. (1986). *The fragility of goodness: Luck and ethics in Greek tragedy and philosophy*. Cambridge: Cambridge University Press.
Nussbaum, M. C. (1994). *The therapy of desire: Theory and practice in Hellenistic ethics*. Princeton, NJ: Princeton University Press.
Oatley, K. (2004). *Emotions: A brief history*. Oxford: Blackwell.
Parkinson, B. (1995). *Ideas and realities of emotions*. London: Routledge.
Picard, R. (1997). *Affective computing*. Cambridge, MA: MIT Press.
Prinz, J. (2004a). Embodied emotions. In R. Solomon (Ed.), *Thinking about feeling* (pp. 44–58). Oxford: Oxford University Press.
Prinz, J. (2004b). *Gut reactions*. Oxford: Oxford University Press.
Robinson, J. (2005). *Deeper than reason*. Oxford: Oxford University Press.
Scherer, K., Schorr, A., & Johnstone, T. (Eds.). (2001). *Appraisal processes in emotion*. Oxford: Oxford University Press.
Shields, S., & Kappas, A. (2006). Magda B. Arnold's contributions to emotion research and theory [Special issue]. *Cognition and Emotion, 20*(7), 898–901.
Sloman, A. (1987). Motives, mechanisms, and emotions. *Cognition and Emotion, 1*, 217–233.
Smith, C., & Lazarus, R. (1993). Appraisal components, core relational themes, and the emotions. *Cognition & Emotion, 7*, 233–269.
Solomon, R. (1984). Emotions and choice. In J. Dougherty (Ed.), *The good life and its pursuit* (pp. 111–132). New York: Paragon House.
Solomon, R. (1988). On emotions as judgments. *American Philosophical Quarterly, 25*, 183–191.
Solomon, R. (1993a). *The passions: Emotions and the meaning of life*. Indianapolis, IN: Hackett Publishing Company.
Solomon, R. (1993b). The philosophy of emotions. In M. Lewis & J. M. Haviland (Eds.), *Handbook of emotions* (pp. 3–15). New York: Guilford Press.
Solomon, R. (2004a). Emotions, thoughts, and feelings: Emotions as engagements with the world. In R. Solomon (Ed.), *Thinking about feeling* (pp. 76–88). Oxford: Oxford University Press.
Solomon, R. (Ed.). (2004b). *Thinking about feeling*. Oxford: Oxford University Press.
Spackman, M. (2004). Can machines adequately simulate human emotion? A test of four theories of emotion. *Theory and Psychology, 14*, 755–776.
Varela, F., Thompson, E., & Rosch, E. (1993). *The embodied mind*. Cambridge, MA: MIT Press.
Wheeler, M. (2005). *Reconstructing the cognitive world*. Cambridge, MA: MIT Press.
Winograd, T., & Flores, F. (1986). *Understanding computers and cognition*. Norwood, NJ: Ablex Publishing Corporation.
Zajonc, R. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist, 35*, 151–175.
Zajonc, R. (1984). On the primacy of affect. *American Psychologist, 39*(1), 117–123.
Zajonc, R., Murphy, S., & Ingelhart, M. (1989). Feeling and facial efference: Implications of the vascular theory of emotion. *Psychological Bulletin, 96*, 395–416.